Standard Operating Procedures for Shoreline Data Development Quality Assurance/Quality Control

The National Oceanic and Atmospheric Administration (NOAA) Coastal Services Center has been creating shoreline data based on National Ocean Service (NOS) shoreline maps for the past three years. Upon the creation of these data, a series of quality checks are performed to assure that the information is as accurate as possible. These standard operating procedures will document the various methods in which the Center checks for accuracy in the shoreline maps that have been vectorized at this office, as well as, in work that has been completed at Techni Graphic Systems (TGS), a NOAA contractor in Ft. Collins, Colorado.

Various methods of quality assurance/quality control are employed and include the following:

1. **Review of line quality during vectorization**
2. **Review of line quality after vectorization**
3. **Review of attribute quality of lines and polygons after vectorization**
4. **Review of line and attribute consistency during edge matching**
5. **Review of attribute table information**
6. **Review of vectorized shoreline compared with established national shoreline**
7. **Review of composite shoreline files by internal review**
8. **Review of composite shoreline files by state partners and the National Geodetic Survey**
9. **Review of metadata**

1. Review of line quality during vectorization

The shoreline vectorization project relies on the use of NOS shoreline maps (sometimes known as t-sheets) that have been scanned and saved in a digital data format. This data format is known as raster data and is comprised of linework that is illustrated as a series of very small squares. The goal of this project is to trace the linework that is drawn on a computer screen to create a different type of digital data format. This process of tracing is known as heads-up digitizing or "vectorizing", whereby the resulting data (known as vector data) are made up of a series of lines resembling those on the raster image. The vector data format is more useful in geographic information systems (GIS).

During the data capture process, the digitizing operator is mindful of the basic principles of how to vectorize accurately. One of the principles is knowing that the traced line must follow directly down the middle of the raster line. Typically, most raster images have linework that is at least four squares, or pixels, across. When the operator is zoomed in on the image close enough, it is possible to see easily if the traced line follows the center of the raster line. Also, with the use of special GIS software, it is possible to allow the vectorization program to automatically trace the lines accurately. However, there are instances when the operator must visually inspect the work of the "auto tracer" to ensure the program is behaving as expected. This visual inspection is a beneficial process that can help catch any problems early on.

2. Review of line quality after vectorization

After the vectorization is complete, the operator creates line quality maps to ensure all necessary shoreline features were captured, and done so accurately. This quality check is accomplished through the production of two types of maps printed on an E-size (34-inch-wide) plotter. The first one is the basemap which is a reproduction of the scanned shoreline map printed on a sheet of opaque bond paper. The second map is printed on translucent vellum paper showing the accomplished linework. This line map is overlaid on a light table on top of the basemap of the raster image. Because the vellum is translucent, the reviewer will be able to see through it and compare it with the lines illustrated on the basemap. The plotter devices that the Center uses are the Hewlitt-Packard 650c or 755 DesignJet Inkjet plotters. The output accuracy of

these plotters allows the resulting two map types to be overlaid one-to-one so that a visual check may occur.

The operator typically is the first one to perform this review, followed by another individual who is more experienced in the light table review. In the case of the work being completed by TGS staff, once they submit their data to the Center, it is reviewed once again by the project manager. Errors that the reviewers look for include instances of linework that "fall off" the raster line, linework that is in need of reshaping by ways of adding or deleting extraneous points, and features that were omitted in the vectorization. These types of errors are circled and brought to the attention of the operator who then goes back and makes the necessary corrections. Another lineplot map may be generated if the reviewer feels as though a significant number of errors were present in the subsequent one. However, if minor changes are needed, then the reviewer may simply request to see the edits made directly to the vector file on the computer in order to save time and money in reproduction of additional maps.

One other aspect that the reviewer must look for is connectivity in the linework. When creating spatial data for use in a GIS, it is essential that all lines connect properly so that there are no gaps, called "undershoots," nor are there lines that cross over too far thus producing "overshoots." To determine if these types of line errors exist, the lineplot must show locations of "dangling" lines. The reviewer must look at examples of these, which are illustrated by the presence of red boxes at the ends of dangling lines and determine if the dangles are actual errors, or possibly the normal and accurate location of a pier or some other man-made feature. Should erroneous dangles be identified, the operator must correct these as well.

3. Review of attribute quality of lines and polygons after vectorization

Once all edits to the linework are complete, the operator will code all of the lines, and in the case of TGS, the polygons as well, with the appropriate feature code used to describe the linear or polygonal features. There is a list of accepted line and polygon feature codes that are used in this process that was developed by a working group of federal and state representatives familiar with the vectorization work. Once this process is complete, the operator generates another map that illustrates the attributes in various colors. This "color map" is then overlaid on the basemap to check for accuracy of labeling lines and polygons. Again, the operator is the first to review, followed by the more experienced analyst, thus providing an opportunity for two sets of eyes to catch any problems.

The vectorized linework is colored so that users can discern the various types of shoreline and man-made features. The polygonal areas are also shaded different colors to show land, water, and man-made features. The reviewers look for instances of erroneous attribution, circle the errors, and bring them to the attention of the operator who then goes back and makes the necessary changes. Another color map may be generated if the reviewer feels as though a significant number of errors were present in the subsequent one. However, if minor changes are needed, then the reviewer may simply request to see the edits made directly to the computer in order to save time and money in reproduction of additional maps.

4. Review of line and attribute consistency during edge matching

Once all of the t-sheets have been finalized, it is necessary to join them together in some fashion so that a large area of data can be displayed in one file. Traditionally, the Center has joined together the most contemporary t-sheets in a given state so that only one file needs to be downloaded from the shoreline Web site. However, the new contract with TGS is set up so that only t-sheets that share photography dates will be appended together. These areas are known as "projects" and can vary in size from 4 to 30 t-sheets.

Prior to appending these files together, another step of attribution is performed. Because a numeric value has been given to each line segment that identifies its feature code, a feature name is assigned so that users can see a text string of information describing the feature. In other words, a typical line segment may have a feature code of "3," but the operator then assigns a feature name of "Apparent Shoreline." Additionally,

other fields are added to the attribute table: t-sheet number, scale, survey date, GIS date. These fields are populated before another process known as edge matching begins.

Edge matching is the method of adjusting the location of linework from one map to meet the location of linework present on adjacent maps. Typically, when t-sheets within a project are edge matched, the process is very simple and the edits are minor. This is because the linework was compiled from the same photographs and crosses over to adjacent maps without any problem. However, the Center has traditionally joined together t-sheets from various projects so that statewide files can be created. In so doing, there is an element of "cartographic license" used to allow continuation of linework. Occasionally, gaps will be present on maps drawn from photography of different dates. When this happens, great care is taken to ensure that the amount of movement of linework is minimal. Should a large gap exist that is too great to resolve from moving the ends of the lines, a phantom line is added and coded as a "User Added Arc." This is done so that the users of the spatial data will be aware that the added segment does not reflect in any way the location of the shoreline, but merely provides an opportunity for connectivity of the linework. TGS staff are finding that the edge matching process can help them identify polygonal areas that were miscoded and make those changes.

5. Review of attribute table information

Once the statewide, or projectwide file is created, a check of its attribute table is necessary. This is done by sorting the fields in ascending or descending order and looking for missing values, values that do not seem logical, or other erroneous input. Errors found are then corrected in the file.

6. Review of vectorized shoreline compared with established national shoreline

A visual check of the statewide or projectwide file is performed by comparing it to the NOAA 1:70,000 median resolution national shoreline file. By doing so, this gives the reviewer another opportunity to verify that the data are in the right spatial location. Any discrepancies are noted, and editing must be performed to ensure an accurate depiction of information.

7. Review of composite shoreline files by internal review

Various staff from the Center will take time to review the data to verify that they appear to look accurate and "make sense." Ideally, staff members who are familiar with a particular state or project area will be able to be involved in the internal review process so that they might lend additional expertise.

8. Review of composite shoreline files by state partners and the National Geodetic Survey

The Center will enlist the help of state partners to review the files, especially for data in their respective states. Members of the National Geodetic Survey (NGS) will also have an opportunity to review and give comments. NGS is the agency in charge of creating the original t-sheets. This will ensure that other sets of qualified eyes will have a chance to review and make comments.

9. Review of metadata

Federal Geographic Data Committee (FGDC)-compliant metadata records are created for each "final" product. In the case of the statewide coverages created at the Center, there is a single metadata written. However, for the new contract with TGS, there are two different types of metadata records. The first is a

record for every scanned t-sheet image, which details the methodology of how the image was originally scanned and georeferenced. The second record is for the edge matched and appended projects. This record details the methodology of how the images were vectorized and then put through the process of quality assurance/quality control. The Center's metadata specialist will assist in checking all of the records to make sure they comply with FGDC standards. Once the spatial data and metadata have been approved as "final" they will be made available eventually on The Center's shoreline data Web site.

The metadata records will also be searchable by the Center's Coastal Information Directory, as well as, by the FGDC clearinghouse. In the near future, all of the spatial data and metadata will migrate over to the National Geodetic Survey, which is currently developing a shoreline delivery Web site using ArcIMS. These data will be available along with other NOS shoreline data that have been created in the past.